

ИССЛЕДОВАНИЕ ЭФФЕКТИВНОСТИ СВЕРТОЧНЫХ НЕЙРОННЫХ СЕТЕЙ КЛАССА U-NET

Н.А. Игольников, Н.Г. Марков
Томский политехнический университет
E-mail: nai6@tpu.ru

Введение

В настоящее время общепринятым и дающим наиболее точные результаты подходом к решению задачи семантической сегментации изображений является применение сверточных нейронных сетей (СНС). Для решения этой задачи широко применяются СНС класса U-Net — архитектура U-Net [1] и производные от нее архитектуры. Однако большая часть работ при реализации такого подхода к решению данной задачи нацелена на достижение высокого качества сегментации и оставляет открытыми вопросы о скорости работы модели, сложности модели и затратах вычислительных ресурсов [2]. Во многих сферах деятельности человека существует необходимость решения задачи семантической сегментации изображений с помощью аппаратных средств, в частности создаваемых на основе программируемых логических интегральных схем (ПЛИС). Для этого необходимо использовать архитектуры СНС с минимальными затратами по памяти и вычислениям. Поэтому задачи исследования эффективности СНС класса U-Net, применяемых для решения проблем семантической сегментации изображений, являются актуальными.

Целью данной работы является анализ эффективности СНС класса U-Net с целью их последующей аппаратной реализации. В перспективе – аппаратная реализация таких СНС на ПЛИС.

Задача исследования эффективности архитектур СНС класса U-Net

В качестве исследуемых архитектур выбраны СНС класса U-Net, описанные в [3]:

- архитектура U-Net;
- разреженная U-Net архитектура.

Архитектура U-Net является архитектурой типа автоэнкодер и состоит из энкодера и декодера. Энкодер и декодер состоят из блоков, включающих сверточные слои, слои нормализации и активации и слои понижения либо повышения размерности. Между энкодером и декодером производятся пробросы карт признаков.

В разреженной архитектуре U-Net используются слои разреженной свертки [4]. Сверточные ядра с различными факторами разреженности показаны на рисунке 1.

Ненулевые весовые коэффициенты ядер на рисунке 1 отмечены точками, все остальные весовые коэффициенты равны нулю. Сверточное ядро с фактором разреженности, равным единице,

эквивалентно ядру обычной свертки. Сверточное ядро размерностью 3×3 с фактором разреженности 2 имеет область восприятия размерностью 5×5 , с фактором 4 — 9×9 .

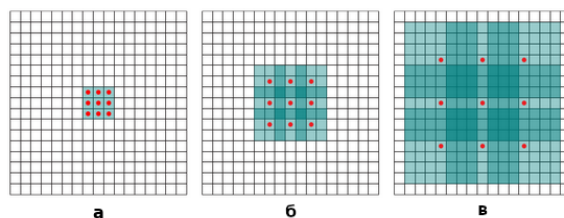


Рис. 1. Сверточные ядра с различным фактором разреженности
а — 1, б — 2, в — 4

Отличие разреженной U-Net архитектуры от обычной состоит в том, что в блоках энкодера и декодера две последовательные свертки 3×3 заменены одной разреженной сверткой 3×3 с фактором 2.

Две последовательные свертки 3×3 эквивалентны одной свертке 5×5 [5]. В первом случае количество весовых коэффициентов равно 18, а во втором — 25. Разреженная свертка 3×3 с фактором 2 также имеет область восприятия 5×5 , а количество весовых коэффициентов в таком ядре свертки равно 9. Таким образом, в ядре разреженной свертки весовых коэффициентов в два раза меньше при неизменной области восприятия. Это, по-видимому, значительно упрощает и ускоряет вычисление U-Net. Исследуем эффективность СНС U-Net и разреженной U-Net.

В качестве обучающей выборки использован датасет TGS Salt Identification [6]. Пример изображения из выборки показан на рисунке 2.

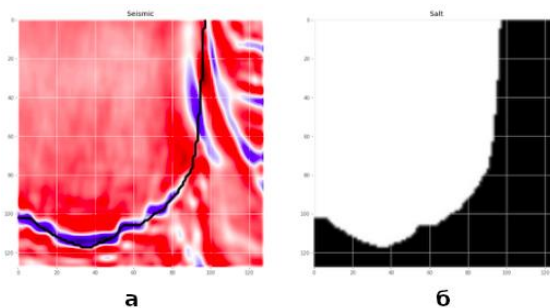


Рис. 2. Пример изображения из обучающей выборки
а — исходное изображение, б — маска сегментации

Размерность изображений выборки — 128x128x1. Обучение моделей проводилось на подмножестве выборки из 3600 изображений. Валидация проводилась на 400 изображениях. Размер батча выбран равным 16, как наиболее успешный компромисс между временем обучения модели и скоростью сходимости процесса обучения. В качестве алгоритма оптимизации выбран алгоритм Adam. Для обучения моделей использовался сервис Google Colaboratory [7].

Сравнение исследуемых архитектур СНС производилось по следующим основным критериям:

- точность сегментации (метрика ассигура — количество верно классифицированных пикселей, деленное на общее число пикселей изображения);
- время прямого прохода изображений;
- количество операций с плавающей запятой;
- количество хранимых чисел с плавающей запятой (объем промежуточных буферов и весовых коэффициентов СНС).

Результаты исследований

Результаты сравнения исследуемых архитектур СНС представлены в таблице.

Таблица. Результаты эффективности архитектур класса U-Net

Критерий сравнения	U-Net	Разреженная U-Net
Точность сегментации изображений обучающей выборки, %	94.7148	91.6562
Точность сегментации изображений валидационной выборки, %	90.0218	89.8101
Время прямого прохода одного батча, мс	2.35	1.99
Умножения, MFLOP	64.98	31.8
Сложения, MFLOP	64.98	31.8
Сравнения, KFL0P	666.62	428.54
Суммарный объем промежуточных буферов, МБ	5.52	3.607
Объем весовых коэффициентов сверточных фильтров, МБ	1.9661	1.0807

Видим, что разреженная U-Net архитектура незначительно уступает в точности сегментации изображений классической архитектуре U-Net, но менее требовательна к ресурсам памяти и вычислительного устройства.

Заключение

Поставлена актуальная задача анализа эффективности СНС класса U-Net с целью их последующей аппаратной реализации.

В ходе исследования вычислительной сложности СНС класса U-Net сделан вывод, что более подходящей архитектурой для аппаратной реализации, например, на ПЛИС является разреженная U-Net архитектура.

Разреженная U-Net архитектура менее требовательна к вычислительным ресурсам при аппаратной реализации и показывает меньшее время прямого прохода сегментируемых изображений.

При этом такая архитектура лишь незначительно уступает в точности сегментации изображений более сложной и медленной архитектуре U-Net.

Благодарность

Исследования были поддержаны грантом РФФИ №18-47-700010р-а.

Список использованных источников

1. Olaf R. U-Net: Convolutional Networks for Biomedical Image Segmentation [Электронный ресурс] URL: arXiv:1505.04597 [cs.CV] (дата обращения: 18.08.2019).
2. Hengshuang Z. ICNet for Real-Time Semantic Segmentation on High-Resolution Images [Электронный ресурс] URL: arXiv:1704.08545 [cs.CV] (дата обращения: 05.07.2019).
3. Данилов В. В. Сегментация и локализация медицинского инструмента при проведении минимально инвазивных операций – Томск, изд-во ТПУ, 2019 – 137 с.
4. Fisher Y. Multi scale context aggregation by dilated convolutions [Электронный ресурс] URL: arXiv:1511.07122 [cs.CV] (дата обращения: 30.08.2019)
5. CS231n Convolutional Neural Networks for Visual Recognition [Электронный ресурс] URL: http://cs231n.github.io/convolutional-networks/#layerpat (дата обращения: 30.08.2019)
6. TGS Salt Identification Challenge [Электронный ресурс] URL: https://www.kaggle.com/c/tgs-salt-identification-challenge/data (дата обращения: 21.08.2019)
7. Google Colaboratory [Электронный ресурс] URL: https://colab.research.google.com (дата обращения: 20.08.2019)